Suitability Analysis of Continuous-Use Reliability Growth Projection Models

THESIS

MARCH 2015

Benjamin R. Mayo, Captain, USAF

AFIT-ENS-MS-15-M-120

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-15-M-120

SUITABILITY ANALYSIS OF CONTINUOUS-USE RELIABILITY GROWTH PROJECTION MODELS

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Benjamin R. Mayo, B.S.

Captain, USAF

MARCH 2015

AFIT-ENS-MS-15-M-120

SUITABILITY ANALYSIS OF CONTINUOUS-USE RELIABILITY GROWTH PROJECTION MODELS

THESIS

Benjamin R. Mayo, B.S.
Captain, USAF

Committee Membership:

Darryl Ahner, PhD, PE
Chair

Major Jason Freels, PhD
Member

AFIT-ENS-MS-15-M-120

Abstract

Substandard system reliability is one of the leading causes of increased Operations and Maintenance (O&M) costs as noted in several recent National Research Council reports. Between 2006 and 2011, Director Operational Test & Evaluation noted 26 of 52 Department of Defense acquisition programs failed to meet reliability thresholds, but were approved, leading to degraded operational performance, increased O&M costs, and increased safety risks for personnel involved. As a system is developed from prototype to final product, structural changes and design flaws are corrected, leading to an increase in system reliability, called reliability growth. Due to the nature of the system changes, standard forecasting methods cannot be applied, and a class of reliability growth models is used to estimate the change in reliability over multiple stages. Despite the significant impact of reliability growth projection, little research has been accomplished on comparing the robustness of various reliability growth models. A simulation is developed to create realistic reliability growth testing data based on historical reliability tests. Using data created via reliability testing simulation, reliability growth projection models are compared based on accuracy and predictive tendencies. Statistical analysis is used to determine which projection models are robust to violations of model assumptions as well as potential hazards in reliability growth program modeling and implementation.

Dedicated to my mother, who fostered a love of math throughout my life, my father, who encouraged a dedication to service and hard work, and my wife, who put up with the many hours of math and hard work, as well as the sacrifices she made while supporting me in this endeavor. You have made this all possible.

Acknowledgements

"The basic process involved is one of learning through failures"

-J.T. Duane

I would like to sincerely thank my advisor, Dr. Darryl Ahner, and co-advisor, Dr. Jason Freels, for the support and belief in my abilities and work. Their guidance and effort helped to shape me not only as an analyst, but as an officer as well. Working with them has fostered my love of analysis that will stay will me throughout my career. Through their patient corrective actions I have grown far more as a researcher, analyst and writer through this last year than I thought possible.

Benjamin R. Mayo

Table of Contents

Page

List of Figures

List of Tables

Table		Page

SUITABILITY ANALYSIS OF CONTINUOUS-USE
RELIABILITY GROWTH PROJECTION MODELS

I. Introduction

Despite all attempts to the contrary, all systems, from simple machines like a pulley to complex, next-generation fighter jets, will break as they are used. Companies creating new products must take this into account, both in the design of the system and the plans for maintenance and repair. Because of this, it is important to determine the probability that a system will function for a given operating time, known as the system reliability[10]. The most common metric for comparing a system's reliability is the Mean Time Between Failures (MTBF), which is the total amount of time the system was operating divided by the number of failures that occurred.

Reliability plays a key role in the operation and maintenance costs of a system. If reliability is overestimated during development, the system may become overburdened with unscheduled maintenance and excess repair costs in the field. Cost studies show that operation and maintenance costs can take up to 84% of a system's life cycle cost [16]. Unfortunately, it is very difficult to estimate a system's ultimate reliability during the early stages of development. In recent years, reliability growth models have gained interest in government acquisition to remove some of the uncertainty from the estimation of reliability.

MIL-HDBK-189C defines reliability growth as "the positive improvement in a reliability parameter over a period of time due to implementation of corrective actions to system design, operation or maintenance procedures, or the associated manufacturing process." [1] This means that for systems in development, reliability improves

as flaws (in reliability growth terms, failure modes) are discovered and fixed. The handbook[1] distinguishes between a repair and a fix (corrective action). A repair is the simple replacement of a part with the exact same components as before the break; in essence, we are still dealing with the same system as before. A fix, on the other hand, is some manner of re-engineering the system into a new, and presumably improved, system [1]. This is one of the reasons that reliability growth is so hard to project: every time failure modes are corrected, the entire system has changed and none of the previous data is valid for extrapolation.

Understanding the concept of failure modes is very important to understanding reliability growth. A failure mode is a design flaw (faulty component or interaction of components) within the system that is believed to be the cause or at least associated with a system failure. In reliability growth literature, failure modes are often classified as either A-modes (failure modes that will not be fixed) or B-modes (failure modes that will be fixed) [1]. In addition to failure modes, reliability growth models use another concept known as the Fix Effectiveness Factor (FEF). This is an assumed percentage reduction in a given failure mode's failure rate based on the fix applied to that failure mode [1]. The FEF plays a key role in growth projection and overestimating it can cause large errors in the model.

Hall [16] notes that reliability growth models can be divided into 3 categories: planning, tracking, and projecting (the same classification is used in [1]). Reliability growth planning deals specifically with the fact that initial system designs and prototypes will have a number of unknown flaws that will prevent the system from achieving the necessary threshold reliability. Reliability growth planning models are used to construct a reliability growth planning curve, which serves to set periodic goals and a benchmark to which the system managers can be held accountable. Comparing the observed system reliability to the planning curve is meant to provide an indication

of the system's progress and earlier indications should problems arise . Assessing the system's actual reliability growth is done with reliability tracking models. These models deal in the area of reliability growth that is most developed and understood[16].

While comparing the reliability tracking data to the reliability planning curve is useful and informative, it can only tell how well a system has progressed so far. Reliability projection models focus on future performance as a function of the current performance, the number of known failure modes, and the fix effectiveness factor. There are two types of projection models: those that assume that the failure modes will not be fixed until after the current test phase is over, and those that allow for some failure modes to be fixed once they are discovered [16].

Reliability growth models have been developed for a variety of systems. From hardware to software, single use to repair and reuse, and discrete use to continuous use, a model exists for all types, shapes, and sizes. The primary focus of this study is a comparison of reliability growth projection models designed for continuous use, repairable hardware systems. These types of models are used to project everything from the battery life of cell phones to the mission capability of next generation aircraft.

In the years since the reliability community first took notice of the Duane reliability growth model, many new reliability growth projection models have been developed and compared to the original. The most popular models are the original Duane model and the Crow model (also known as the AMSAA model). While most authors compare new models to either the Duane or the Crow models, based on the research in this paper, a comparison of multiple models against realistic reliability growth data is unprecedented. To that end, this research compares 9 continuous-use reliability growth projection models against simulated failure times in order to determine which models are most appropriate for what types of reliability testing, as well as how robust these models are to violations of their assumptions and constraints.

II. Literature Review

2.1 Reliability

Reliability is commonly defined as "the probability that a system, vehicle, machine, device, and so on will perform its intended function under operating conditions, for a specified period of time" [18]. The most common metric used for measuring reliability in repairable systems is the Mean Time Between Failures (MTBF), also known as the Mean Time Between Critical Failures, Mean Time Between Operational Failures [19]:

$$MTBF = \frac{\sum_{i=1}^{n} t_i}{n} \tag{1}$$

where n is the number of failures and t_i is the operational time between failure $i-1$ and failure i, with $t_0 = 0$. Another alternative measure is the average cumulative number of failures at time T, the total operational testing time[19].

2.2 Reliability Growth

The concept of reliability growth has been a focus of development since manufacturing began, but it was not until the 1950's that growth potential was first modeled [22]. In recent years, reliability growth models have come to the attention of both government and commercial agencies. In 2002, the National Research Council (NRC) conducted a workshop on Reliability for DoD Systems, outlining the history of reliability growth modeling and advocating for their use on developing DoD systems. The workshop determined that the use of reliability growth models along with reliability-focused system design had the potential to prevent cost overruns in new DoD systems [3]. In addition to [3], the DoD continues to update the Handbook for Reliability Growth Management with the latest policy and processes for reliability

growth models [1] to promote the use of reliability growth modeling and management in acquisition systems.

Despite requirements to use reliability growth models, recent studies have noted trends in reliability failures throughout the DoD. In [14] Dr. Michael Gilmore (Director, Operational Test & Evaluation) noted that since 1985, 51 of 170 DoD systems failed to meet reliability requirements. In 2014, the Panel on Reliability Growth Methods for Defense Systems for the NRC published a second report that showed 26 of 52 major Department of Defense programs failed to meet the reliability goals set for them between 2006 and 2011 [19]. As discussed in the report, all 52 systems were approved out of necessity. In fact, [19] provides evidence for increased reliability failures throughout the DoD, suggesting that acquisition programs require a more rigorous design for reliability in their testing. Fielding these programs would lead to significantly increased maintenance costs and risks to personnel, forcing decision makers to determine what was more costly: approving the program or canceling production and working without.

In addition to the need for reliability design, [19] alludes to issues within many reliability growth models that need to be considered when choosing the model. Many models have assumptions about the improvement process that the system test follows. Some models assume that corrective actions are implemented using a Test-Fix-Test process: systems are tested until a failure occurs, at which point the cause of the failure is determined and corrected, allowing testing to continue. The most common practice currently is the Test-Find-Test process: systems are tested until a failure occurs, the cause of the failure is determined but not corrected until the end of the current testing phase [1]. The 2014 NRC report notes that some models are used on a Test-Find-Test system test despite the model assumption that the test is conducted according to Test-Fix-Test processes. Additionally, many models assume that

non-corrective repairs return the system to “Good-as-New” status, assuming that all failure modes are independent of each other. Most significantly, however, is the recommendation that reliability growth projection models require extrapolation and should not be used for reliability growth estimation without validation [19]

2.3 Reliability Growth Models

2.3.1 Weiss Model.

Weiss [22] discusses reliability growth under Test-Fix-Test conditions: corrections are made as failures occur and the improved system then continues the test until the next failure. Assuming that the failure rates follow a Poisson distribution, Weiss used maximum likelihood estimation to develop a model for the MTBF ($T(i)$) based on the trial number i.

$$T(i) = Ae^{ci} \tag{2}$$

where A and c are parameters determined through maximum likelihood estimation from initial tests[22].

2.3.2 Duane Model.

In 1964 Duane discovered that as a system undergoes design changes to remove failures and grow reliability, plotting the cumulative failure rate against the cumulative test time on log-log scale results in a linear relationship (known as the ”Duane Postulate”)[9]. Initially a graphical representation, Duane regression model for this log-log relationship was

$$\lambda(T) = K(T)^{-\alpha} \tag{3}$$

where α is the log-slope and K is a constant, both determined by regression. $\lambda(T)$ is the cumulative MTBF at time T. Using this equation, Duane was able to estimate

the cumulative number of failures, the average failure rate, and the instantaneous failure rate. Duane's Postulate was developed after observing the cumulative data for aircraft of varying size and complexity.[1] [9]. While Duane was not the first to develop a reliability growth model, his model became the basis for future reliability growth models for decades.

2.3.3 AMSAA Crow Model.

Larry H. Crow published a paper in 1975 in which he develops a model, incorporating methods from both the Duane Model and the Weiss model. Crow states that if the Duane Postulate is correct, then the failure rate follows a nonhomogeneous Poisson process, with a Weibull intensity function [4]

$$u(T) = \lambda\beta T^{\beta-1} \tag{4}$$

where λ and β are determined by maximum likelihood estimation. As long as $0 < \beta < 1$, the system reliability is increasing[4]. Assuming a Test-Fix-Test strategy, Crow shows that the failure rate should decrease over time as more failure modes are discovered and fixed. Using the intensity function, Crow developed an equation for the probability of failure ($f(T)$) during a fixed time interval (d) [4]:

$$f(T) = e^{-[\lambda(T+d)^{\beta} - \lambda(T)^{\beta}]} \tag{5}$$

2.4 AMSAA-Crow Projection Model

Crow [5] expands upon [4] to incorporate the idea that not all failure modes within the system will be fixed. Designating the failures that will remain unchanged throughout the testing procedure as Type A failures and corrected failures as Type B failures, Crow also described the idea of a Fix-Effectiveness Factor (FEF) that would

capture how the corrective actions would effect the failure rates of the Type B failure modes [5]. This allowed for the possibility of projecting the new failure rate into the next series of tests or phases. The new failure rate ($f(T)$) becomes[5]

$$f(T) = \lambda_A + \lambda_B - \sum_{i=1}^{M} d_i \lambda_i \tag{6}$$

where λ_A and λ_B denote the failure rates of the Type A and corrected Type B failure modes, respectively, λ_i denotes the original failure rate of the ith corrected Type B failure mode (for M corrections), and d_i denotes the FEF for the ith corrected failure mode. This method also allows for a Test-Find-Test strategy, where the corrective actions can be delayed until the end of a testing phase, allowing for longer test times and potentially greater improvement in the reliability parameters [5]. To estimate the growth, $\rho(T)$, Crow [5] developed the following equation for the change in failure intensity rate:

$$\rho(t) = \lambda_A + \sum_{i=1}^{M}(1 - d_i)\lambda_i + \mu_d h_c(t) \tag{7}$$

where μ_d is the average FEF over all discovered FMs, and $h_c(t)$ is the expected number of new failure modes that are discovered in the next time interval derived as:

$$h_c(t) = \lambda\beta t^{\beta-1} \tag{8}$$

However, as the true λ and β are unknown, they are estimated with $\hat{h}_c(t)$[1]

$$\hat{h}_c(t) = \frac{m\hat{\beta}}{T} \tag{9}$$

$$\hat{\beta} = \frac{m}{\sum_{i=1}^{m} \ln\left(\frac{T}{t_i}\right)} \tag{10}$$

where t_i is the time of failure i[1].

2.5 Crow Extended Reliability Projection Model

After Crow developed models around delayed and uncorrected fixes, he developed the Crow Extended Projection Model to incorporate corrections during a phase, allowing for a test that combines Test-Find-Test and Test-Fix-Test methods. Crow designated BC failure modes as those that are corrected during the testing phase and BD failure modes as those that are corrected at the end of the phase [1][6]. The Extended Projection equation is

$$\lambda = \lambda_{CA} - \lambda_{BD} + \sum_{i=1}^{M}(1 - d_i)\lambda_i + \mu_d h(T|BD) \tag{11}$$

λ_{CA} is the current estimated failure rate, typically gathered from the Crow Tracking model. The remaining terms are calculated the same way as the AMSAA-Crow Projection Model with respect to the BD failure modes, thus if no corrections are made during the phase, the extended model becomes the same as the AMSAA-Crow Projection Model [1].

2.6 Variance-Stabilized Duane

The Duane model is often criticized for violating many of the assumptions required for simple linear regression, specifically that the variance is constant and normally distributed [7]. Donovan and Murphy [8] developed a new regression model based on variance stabilization techniques that follows the same system assumptions as the Duane (earning it the nickname, Variance-Stabilized Duane Model). This model places more influence on the most recent failures, suggesting that they have more information about the failure rate from the next phase [8]:

$$\theta = \alpha + \beta\sqrt{T} \tag{12}$$

Due to the similar forms, if the slope of the Duane Model is 0.5, the authors note that the Variance-Stabilized Duane and the Duane model are mathematically equivalent [8].

2.7 AMSAA Maturity Projection Model

Ellner [12] developed the AMSAA Parametric Empirical Bayes projection model (now known as the AMSAA Maturity Projection Model). This model allows for Type A, Type BC, and Type BD failure modes and estimates the discovery rate of new Type B failure modes[12]. This model assumes that the failure rates for each failure mode, λ_i are random samples from a random variable that follows the gamma distribution, $\Gamma(\alpha, \beta)$. By estimating the true failure rates, Λ_i, from the observed λ_i, the Maturity Projection Model estimates the failure intensity[1]:

$$\rho(t : \Lambda) = \lambda_A + \sum_{i=1}^{m}(1 - d_i)\Lambda_i + \sum_{i=1}^{m} d_i \Lambda_i e^{-\Lambda_i t} \tag{13}$$

the expected value of which is

$$\rho(t) = \lambda_A + (1 - \mu_d)\lambda_K + \mu_d h(t) \tag{14}$$

In order to find the components of the equation (based on the K discovered failure modes), Ellner describes the MLE for β_k -$\hat{\beta}_k$ as

$$K = \frac{\left(\sum_{i=1}^{m} \ln \frac{1+\hat{\beta}_k T}{1+\hat{\beta}_k t_i}\right)\left(\sum_{i=1}^{m} \frac{1}{1+\hat{\beta}_k t_i}\right) - \left(\frac{m\hat{\beta}_k}{1+\hat{\beta}_k T}\right)\sum_{i=1}^{m}\frac{T-t_i}{1+\hat{\beta}_k t_i}}{\ln 1 + \hat{\beta}_k T\left(\sum_{i=1}^{m}\frac{1}{1+\hat{\beta}_k t_i}\right) - \left(\frac{m\hat{\beta}_k}{1+\hat{\beta}_k T}\right)T} \tag{15}$$

And $\hat{\alpha}_k$ is found by

$$(\alpha_k + 1)^{-1} = m^{-1} \left[K \ln 1 + \hat{\beta}_k T - \sum_{i=1}^{m} \ln \frac{1 + \hat{\beta}_k T}{1 + \hat{\beta}_k t_i} \right] \tag{16}$$

Using the determined parameter values from 15 and 16, the following equations are used for inputs in equation 14[1]:

$$\lambda_i = \frac{m\hat{\beta}_i}{\ln 1 + \hat{\beta}_i T} \tag{17}$$

$$\lambda_A = \frac{N_A}{T} \tag{18}$$

$$\mu_d = \frac{1}{m} \sum_{i=1}^{m} d_i \tag{19}$$

$$\hat{h}(t) = \frac{\lambda_i}{1 + \hat{\beta}_i t} \tag{20}$$

2.8 AMSAA Maturity Projection Model - Stein

In 2004, Ellner [11] published a variant of the AMSAA Maturity Projection Model by incorporating the Stein Estimation process [20]. This process provides an estimate of the individual failure rates (λ_i):

$$\tilde{\lambda}_i = \theta \hat{\lambda}_i + (1 - \theta) \left(\frac{1}{k} \sum_{i=1}^{k} \hat{\lambda}_i \right) \tag{21}$$

where $\theta \in [0, 1]$ is the value that generates the minimum sum of squares $\sum_{i=1}^{k} (\tilde{\lambda}_i - \lambda_i)^2$. The growth potential estimation is then

$$\hat{\rho}(T) = \hat{\lambda}_A + \sum_{i \in obs(B)} (1 - d_i) \tilde{\lambda}_i + \sum_{i \in obs(B)} \tilde{\lambda}_i \tag{22}$$

Ellner derives the value for θ_S as

$$\theta_S = \frac{kVar(\lambda_i)}{kVar(\lambda_i) + \left(\frac{\lambda}{T}\right)\left(1 - \frac{1}{k}\right)} \tag{23}$$

Because θ_S relies on the unknowns k, λ, and $Var(\lambda_i)$, an estimate, $\hat{\theta}_S$, can be estimated via maximum likelihood estimation for a finite number of failure modes k, $\hat{\theta}_{S,K}$ [1] with

$$\lim_{k\to\infty} \hat{\theta}_{S,K} = \hat{\theta}_{S,\infty} = \frac{\hat{\beta}_\infty}{1 + \hat{\beta}_\infty T} \tag{24}$$

$\hat{\beta}_\infty$ is found such that it satisfies

$$m = \left(\frac{N_B}{\hat{\beta}_\infty}\right) \ln(1 + \hat{\beta}_\infty T) \tag{25}$$

where N_B is the number of discovered Type B failures and m is the number of observed Type B failure modes[1]. $\hat{\theta}_{S,\infty}$ is substituted into Equation 21.

2.9 Clark Model

In 1999, Jeffrey A. Clark[2] created a model specifically for later in development when there are fewer failure modes present and it is possible to have eliminated failure modes. For this model, failure modes can be correctable or inherent (Type B or Type A, respectively), but some failure modes can be can be completely corrected (no longer affect the system). The model is

$$\lambda_T = \lambda_I + \lambda_F - d\lambda_{SF} - \lambda_{VF} + \lambda_U \tag{26}$$

λ_T is the projected failure rate

λ_I is the inherent failure rate (Type A failures)

λ_F is the failure rate of correctable failure modes

λ_{SF} is the failure rate of the failure modes that have scheduled corrections

λ_{VF} is the failure rate of failure modes that have been eliminated

λ_U is the failure rate of undiscovered failure modes

d is the average FEF across all correctable failure modes

Clark notes that the individual rates are calculated in the same manner as in the AMSAA Crow Projection Model; however, the Clark model has the added classification of the eliminated failure modes and assumes that the undiscovered failure mode rate ($h(T)$ from Equation 7) is negligible due to testing in later development stages, assuming the phase length is short [2].

2.10 Guo-Zhao Model

A common assumption in reliability growth models is that the intermediate repairs return the system to the pre-breakdown state, but do not affect the failure rate in any way. In 2006, Huairui Guo et al developed a model that allows for an estimate of the repair effects[15]:

$$\lambda(t) = \lambda\beta t^{\beta-1}e^{\gamma N(t)} \tag{27}$$

with $N(t)$ the number of failures by time t and γ the repair effect, and λ, β are model parameters. If $\gamma < 0$, the repairs are increasing the failure rate, $\gamma > 0$ the repairs are decreasing the failure rate and when $\gamma = 0$, the model is the same form as the AMSAA Crow Plannning model.

2.11 Models and Assumptions

Table 1 contains a list of the models considered in this study and their assumptions.

Table 1. Projection Model Assumptions

Model	**Assumptions**
Weiss (1956)	• Failure modes are independent from each other • Failure times are exponentially distributed • Failure rates always decrease • Corrective Actions do not increase the failure rate • High probability of failure means high probability of detection/correction • Reliability Testing occurs during normal operating conditions • Test follows Test-Fix-Test pattern
Duane (1964)	• Failure modes are independent from each other • Failure times are exponentially distributed • Failure rates always decrease • Corrective Actions do not increase the failure rate • High probability of failure means high probability of detection/correction • Reliability Testing occurs during normal operating conditions • Test follows Test-Fix-Test pattern

Crow Projection Model (1984)	• Failure modes are independent from each other • Failure rates follow non-homogeneous Poisson distribution • Test follows a Test-Find-Test pattern • Corrective Actions do not increase the failure rate • High probability of failure means high probability of detection/correction • Reliability Testing occurs during normal operating conditions • Not all Failure Modes must be corrected
Crow Extended Projection Model (2004)	• Failure modes are independent from each other • Failure rates follow non-homogeneous Poisson distribution • Test follows a Test-Find-Test pattern or Test-Fix-Test • Corrective Actions do not increase the failure rate • High probability of failure means high probability of detection/correction • Reliability Testing occurs during normal operating conditions • Not all Failure Modes must be corrected
Variance-Stabilized Duane (2000)	• Failure modes are independent from each other • Failure times are exponentially distributed • Failure rates always decrease • Corrective Actions do not increase the failure rate • High probability of failure means high probability of detection/correction • Reliability Testing occurs during normal operating conditions • Test follows TFT pattern

Maturity Projection Model (1995)	• Failure modes are independent from each other • Test follows a Test-Find-Test pattern or a Test-Fix-Test pattern • High probability of failure means high probability of detection/correction • Initial Type B failure mode failure rates can be modeled as a random sample from a gamma distribution • Not all Failure Modes must be corrected
Maturity Projection Model-Stein (1995)	• Failure modes are independent from each other • Test follows a Test-Find-Test pattern only • High probability of failure means high probability of detection/correction • Initial Type B failure mode failure rates can be modeled as a random sample from a gamma distribution • Not all Failure Modes must be corrected
Clark (1999)	• Failure modes are independent from each other • Failure rates follow non-homogeneous Poisson distribution • Test follows a Test-Find-Test pattern or Test-Fix-Test • Corrective Actions do not increase the failure rate • High probability of failure means high probability of detection/correction • Reliability Testing occurs during normal operating conditions • Not all Failure Modes must be corrected • The testing phase is sufficiently short to assume that no new failure modes are discovered after the first phase

Guo et al (2006)	• Failure modes are independent from each other • Failure rates follow non-homogeneous Poisson distribution • Test follows a Test-Find-Test pattern or Test-Fix-Test • Corrective Actions do not increase the failure rate • Intermediate Repairs can affect the failure rate • High probability of failure means high probability of detection/correction • Reliability Testing occurs during normal operating conditions • Not all Failure Modes must be corrected

III. Methodology

3.1 Research Goals

As previously stated, the goal of this research is to compare modern and historical reliability growth models in their projection performance against simulated reliability testing data. While reliability growth testing has been incorporated into many systems in more recent years, there was no way to guarantee that the model assumptions would be met due to unknowns in testing. To that end, a series of datasets were developed using a simulation in R. The simulation was developed with the goal to replicate the process of contemporary reliability growth testing in order to determine the robustness of each model to violations in accepted assumptions.

3.2 Simulation

The simulation was based on the concept that a system has an inherent (and unknown) number of failure modes at the beginning of the test. Each of these failure modes has an underlying (and also unknown) distribution that can only be discovered when that mode causes a failure. In order to test the accuracy of the models in systems that meet and do not meet the assumptions, failure mode distributions followed either the exponential distribution or the Weibull distribution. A flowchart of the simulation is shown as Figure 1 while a summary of the simulation steps is below:

1. The number of failure modes, types of failure mode distribution, the total test time (in hours), and the number of Corrective Action Periods are given as inputs

2. Failure mode distributions parameters are generated based on the total test time

3. Failure times are generated by randomly sampling from the generated distributions

4. Fix effectiveness factors are generated according to a uniform distribution and applied to the distribution parameters

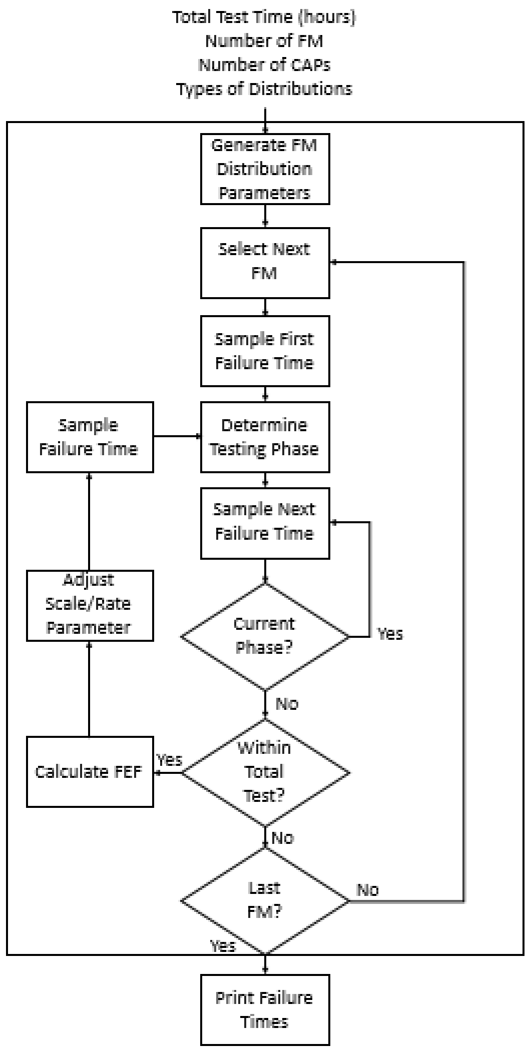


Figure 1. Model Flowchart

3.2.1 Model Inputs.

Each model run is based on the number of failure modes, their distributions, the total amount of time until the test is over, and the number of corrective action periods

(CAPs) which determine the number of testing phases. In order to compare the effects of these inputs, each one was treated as a 3-level factor. The number of failure modes was set at 4, 20, and 36, the distributions were either exponential or Weibull, and the number of CAPs was 2, 5, or 8. The total test time was held constant at 2000 hours.

3.2.2 Failure Mode Distributions.

Each failure mode was based on a specific distribution with its own parameters. The exponential and Weibull distributions are generally used to simulate the time between failures for reliability models [17]. While the Gamma distribution is sometimes considered, it was left out of this simulation due to the fact that the Weibull distribution can approximate the Gamma fairly easily. Many of the models assume that the failure times are distributed exponentially, so a strict exponential distribution was used to stay within their assumptions. In reality, however, reliability growth models often must be used despite failure times that may violate their assumptions. To account for this, the Weibull distribution was also used in order to create "messy" failure distributions in order to test the models' performance against systems that violate assumptions.

For the k exponential distributions, λ_k was generated according to the following equation:

$$\lambda_k = \frac{Unif(0.5, 15)}{T} \tag{28}$$

where T is the total test time (in the case of this study, 2000 hours). The uniform distribution was used to determine the value of the numerator, essentially creating an average time between failures between 133 and 4000 hours. This provides a failure rate that is not so high that it does not occur during the testing time but also that is high enough to avoid failing so often that the failure mode dominates the testing time. The range of the uniform distribution was skewed so that a greater number of failure

modes would appear earlier in the testing cycle, while still allowing for failure modes that would not appear until later. This follows many of the model assumptions that failure modes with greater failure rates are discovered and corrected sooner, allowing for the discovery of failure modes with lower failure rates.

Similar to the exponential distribution, the Weibull distribution has a scale factor (β_k) that determines the failure rate. This was generated according to the following equation:

$$\beta_k = \frac{T}{Unif(0.5, 4)} \tag{29}$$

The range was again skewed towards a higher rate in order to create more failure modes that will occur more often earlier in the cycle. The range of the Weibull scale parameter differs from that of the exponential due to the Weibull's shape parameter. While the scale parameter was based on the total test time in order to avoid failure modes that never occur, the shape parameter (η_k) is produced independently of the total test time:

$$\eta_k = Unif(0.5, 5) \tag{30}$$

This was due to the fact that, regardless of the scale parameter value, the PDF of a Weibull distribution with $\beta < 0.5$ is significantly skewed to the left. Similarly, for $\beta > 5$, the PDF is skewed to the right. In order to prevent failure modes that occur so frequently they skewed the initial reliability too low as well as failure modes that never occur during the testing time, those limits on the scale and shape parameters were deemed the most appropriate.

3.2.3 Failure Times.

Each test has n Corrective Action Periods, giving it $n+1$ testing phases. For each failure mode, an initial failure time is sampled according to its distribution, and the testing phase (p_k) that this failure would occur is recorded. This way, a failure mode that has not occurred in the "test" has no corrective actions applied until it appears. Subsequent failure times are recorded until the the time in phase p_k is realized. At this point, the Fix Effectiveness Factor is applied and a new distribution parameter is assigned. The next failure time is generated, the test phase (p_{k+1}) is determined and the process continues until the Total Test Time has been reached.

3.2.4 Fix Effectiveness Factor.

The Fix Effectiveness Factors (FEF) for each failure mode are only generated once the failure mode has occurred and the testing phase is over. Because the true FEF cannot be determined, many models make use of an average FEF. In order to avoid skewing the results due to the true FEF being much higher or lower than the average, the FEF is generated from the uniform distribution with a minimum of 0.2, maximum of 0.8, which has an average of 0.5. As first suggested by Crow in [5], FEF in later phases are based on the average of the FEF from earlier phases. For the purposes of this simulation, all average FEFs are assumed to be 0.5 for the use of the models, allowing for a simplification of the calculations for the FEF in some models like the Crow Projection [5]. For failure modes with the exponential distribution, the FEF is applied to λ:

$$\lambda_{New} = \lambda_{Old} * FEF \tag{31}$$

This results in the subsequent failure rate being lower than the original. This models the results of an actual corrective action by accounting for the varying efficacies of the redesigning the system and the unknown effects that it will have on that

failure mode in the future. For the Weibull distribution, it is commonly assumed in reliability literature (as noted in [13]) that the shape parameter remains constant and the corrective actions will only affect the scale, β:

$$\beta_{NEW} = \beta_{OLD}/FEF \tag{32}$$

This ensures that the nature of the failure mode (assumed to be the shape parameter of the distribution) remains the same while the rate of occurrence is decreased by the corrective action. For example, if the failure mode is based around a certain component overheating, this model assumes that whatever corrective action is applied only effects how quickly the component breaks down due to heat, not the fact that heat is the overarching design flaw.

3.3 Experiment Design

Each simulation run creates a series of failure times that could occur with the given parameters (number of FMs, CAPs, and the types of distributions). Because the failure rates were created relative to the total test time (see equations 29, 28), the total test time was kept constant for all simulations, allowing for a comparison of shorter and longer testing periods (the result of more and fewer CAPs, respectively). The factors in the design were the number of FMs, the number of CAPs, and the types of distributions for the failure times.

The level settings were determined from AMSAA sample reliability growth data. The sample data contained example reliability growth tests for 12 systems, ranging from radios to air defense systems. The number of FMs and CAPs varied for each example, and the maximum and minimum values were used to determine the high and low levels. For the number of FMs and CAPs, three levels were chosen for testing.

For number of FMs, the low, middle, and high levels were 4, 20, and 36. For the number of CAPs, the low, middle, and high levels were 2, 5, and 8. As noted in [17], the most common parametric distributions for modeling failure times are the exponential and Weibull distributions, therefore these were chosen as the two levels for the types of distribution. Each simulation was replicated three times, resulting in 54 datasets. Table 2 shows the full factorial design for a single replication.

Table 2. Single Replication of Simulation Runs

Run	Failure Modes	Corrective Action Periods	Distribution
1	4	2	Exponential
2	20	2	Exponential
3	36	2	Exponential
4	4	5	Exponential
5	20	5	Exponential
6	36	5	Exponential
7	4	8	Exponential
8	20	8	Exponential
9	36	8	Exponential
10	4	2	Weibull
11	20	2	Weibull
12	36	2	Weibull
13	4	5	Weibull
14	20	5	Weibull
15	36	5	Weibull
16	4	8	Weibull
17	20	8	Weibull
18	36	8	Weibull

The purpose behind this design was to create a series of datasets that mimic the data that is captured during real-world reliability growth testing (like those found in the AMSAA sample data). Modern testing follows the Test-Find-Test corrective action implementation strategy, delaying corrective actions until the end of a given testing period. This allows for more FMs to be discovered and corrected at a given

time. As testing continues, new FMs are discovered and corrected over time, theoretically decreasing the failure rate as they are corrected. This is consistent with the assumptions of Weiss [22] and Duane [9] that FMs with higher failure rates are more likely to be discovered and corrected. To be as similar to contemporary testing as possible, each simulation run creates a series of failures and corrective actions for a system based on the input parameters.

The levels for the design factors were chosen to recreate various testing conditions that can occur for different systems. Datasets with only 4 FMs simulate more mature systems where many early flaws have been discovered and eliminated or simpler systems, while datasets with 36 FMs simulate systems that are early in development with a significant number of flaws and more complex. The number of CAPs were varied to compare how the models perform with varying test period lengths. Most models assume that failures occur according to a Non-Homogeneous Poisson Process (NHPP). To account for this, the failure times were sampled from the exponential distribution with the rate changing as corrective actions are taken. Additionally, datasets were developed with failure times following a Weibull distribution in order to compare the model performance when the NHPP assumptions are violated.

3.3.1 Model Assumptions.

Below is the list of assumptions made during the development of the simulation.

- Failure Modes occur independently from one another
- No new Failure Modes are introduced by corrective action
- Corrective action affects only the Failure Mode to which it is applied
- Corrective action occurs at the end of the testing phase
- All Failure Modes are correctable (Type B)

- Failure times occur according to either an Exponential or Weibull distribution
- Failures cannot be corrected unless they are observed during a specific testing phase
- Intermediate repairs do not affect the failure rate

3.4 Examples

Next, two examples are presented following the methodology in Figure 1. Each example's purpose is to illustrate the methodology and will go through the loop for a single failure mode in a system with the following inputs:

- Total Test Time: 2000 hours
- Number of Failure Modes: 15
- Number of Corrective Action Periods: 2
- Types of Distributions: Exponential

3.4.1 Exponential Distribution.

For this example, failure modes are exponentially distributed.
The distribution for the first failure mode (FM_1) is designated as exponential, so the scale parameter λ_1 is sampled from Equation 28: $\lambda_1 = 0.0055$. There are 2 Corrective Action Periods, meaning that there are 2 Test Phases, each 1000 hours of test time. The first time (T_1) sampled from FM_1 56.80086, which is in Test Phase 1. The current phase is set to 1, the failure time and failure mode are recorded, and the current test time is updated to 56.80086.

The next failure time (T_2) is 23.39892, so the next failure for FM_1 would occur at $T_1 + T_2 = 80.19978$ hours of testing time. This is still within the time for Test Phase

1, so the failure time and mode are recorded and the current test time is updated to 80.19978. This continues until a subsequent failure time occurs at a time greater than 1000 hours, which happens with T_5. The 4th failure occurred at 800.1686 testing hours, and $T_5 = 315.6473$, which would occur after test phase 1 is over. In this case, the failure is not recorded (as it would not have been observed), and the current Test Time is updated to the beginning of test phase 2: 1000 test hours. Because FM_1 was discovered, a corrective action takes place. This is modeled via the FEF. A FEF is calculated as 0.342 and applied to the scale parameter, making the new scale parameter for FM_1 0.001881.

Because a corrective action took place after test phase 1, the next failure time for FM_1 will take place after the beginning of test phase 2. When sampled from the new distribution, T_1 is 631.612 testing hours, meaning the failure occurs at 1631.612 testing hours. Test phase 2 does not end until 2000 testing hours, so the current phase is set to 2, the failure time and failure mode are recorded, and the current test time is set to 1631.612 testing hours.

T_2 occurs after 700.448 testing hours and testing time 2332.060, falling after the end of Test Phase 2 so it is not observed. Again the FEF is generated, this time FEF = 0.7076. The new scale parameter becomes 0.001331 and the model then iterates to the next failure mode, FM_2 and resets the current test time to 0.

3.4.2 Weibull Distribution.

FM_2 is designated as Weibull, so the shape and scale parameter are sampled from Equations 29 and 30 respectively: $\eta = 0.00175$ and $\beta = 2.689$ The first failure (T_1) occurs after 590.8067 hours, and because the current test time is 0, this is within test phase 1, so the current phase is set to 1. The failure mode and time are recorded, and the current test time is updated to 590.8067 hours. The second failure (T_2)

occurs after 723.4597 hours and at testing time 978.4332. Because this would have occurred after test phase 1 had ended, the failure is not recorded, the current test time is updated to 1000, and the FEF is calculated to be 0.2117. For the Weibull, the FEF only applies to the rate parameter and not the shape parameter as mentioned previously. The new rate parameter is 0.00037. The next failure occurs after 1593.859 hours, which is after the Total Test Time is reached, so FM_2 never occurs again during the test cycle. The model would then iterate through the remaining 13 failure modes before completing.

At the end of each test simulation, the model runs for an additional 2000 hours in order to get an estimate of the MTBF after the final CAP. This MTBF is used for the final prediction comparison.

IV. Analysis

4.1 Model Implementation

The simulation runs were generated to test each model's performance against datasets that followed and violated the model's assumptions. However, as the models were implemented, it became apparent that, due to the assumptions of the simulation, some of the projection models would become mathematically equivalent:

- Under the assumption that all corrective actions are delayed, the ACPM-Extended Model is equivalent to the ACPM
- If the Duane and Weiss models are implemented via regression, they become mathematically equivalent
- Under the assumption that repairs have no effect on the reliability of the system, the Guo-Zhao model becomes the Crow Model
- Under the assumption that no FMs can be completely eliminated, the Clark model and the ACPM are identical, provided the $h(T)$ term is not considered negligible; in order to differentiate between the two models, this assumption was made for the Clark Model

With that, the models that were compared were the Duane, AMSAA-Crow, Variance-Stabilized Duane, AMSAA-Maturity Projection, AMSAA-Maturity Projection-Stein, and the Clark Models.

4.2 Run Responses

Each run of the simulation produced a series of failures, separated by phase times. This allows for the estimation of the MTBF for each phase, which is compared to the

projected MTBF for that phase. Example output from the simulation is in Table 3: For n phases, there are $\frac{n(n-1)}{2}$ projections that can be made: Phase 1 can provide a

Table 3. Example Output

Phase	Failures	MTBF
1	150	3.226
2	115	4.348
3	76	6.579
4	40	12.5

projection onto Phases 2, 3, 4, and the end of test; Phase 2 can provide a projection onto Phases 3, 4, and the end of test, and so on. Each projection may have an associated error when compared to the observed MTBF, resulting in $\frac{n(n-1)}{2}$ error calculations from phase i to phase j: $E_{i,j}$. While the most common prediction error calculation is the least squares, $\sum(\hat{y} - y)^2$, the varying number of phases, along with the wide range of MTBF, an attempt to "normalize" the errors is made.

$$E_{i,j} = \frac{(\widehat{MTBF}(i,j) - MTBF(j))^2}{MTBF(j)} \tag{33}$$

where $E_{i,j}$ is the error of the projection from phase i to phase j, $\widehat{MTBF}(i,j)$ is the projected MTBF from phase i onto phase j, and $MTBF(j)$ is the observed MTBF in phase j. This research focuses on the next phase projection error, $E_{i,i+1}$,the final projection error from the initial phase, $(E_{1,n})$, and the projection error from the final phase into the end of testing, $E_{n-1,n}$. In addition to those projection errors, the average projection error across N projections (E_{AVG}) and the maximum projection error (E_{MAX}) are considered.

$$E_{AVG} = \left(\frac{1}{n}\right)\left(\sum_{i=1}^{n} E_{i,i+1}\right) \tag{34}$$

$$E_{MAX} = Max(E_{i,i+1} | 1 \leq i \leq n-1) \tag{35}$$

Some models, such as the Duane and Variance Stabilized Duane models, require at least two points of data in order to make an estimate [22][9][8]. This often requires the use of an estimated MTBF at the beginning of the test. This is usually gathered from previous testing, data from a similar system, or simulation. To account for the impact that this estimate will have, models that require the earlier estimate were run against data with an additional Phase 0, with with the MTBF estimate varied from $0.5 \times (MTBF(1))$ to $1.5 \times (MTBF(1))$ in increments of 0.25. This helps to simulate over and underestimation of the initial MTBF. To test these models without the bias of the initial estimated MTBF, they were also tested against the situation where they could only begin projection in Phase 2 after 2 data points without a Phase 0 being considered.

4.3 Model Comparisons

4.3.1 Projection Proportions.

In order to determine a model's tendency to over or under-predict the MTBF, the proportion of single-phase projection errors ($\hat{p}_{i+1}$) and end-phase projection errors ($\hat{p}_n$) that were negative or positive were calculated.

$$\hat{p}_{i+1}(NEG) = \left(\frac{1}{n}\right) \sum_{\tilde{E}_{i,i+1}<0} 1 \tag{36}$$

$$\hat{p}_{i+1}(POS) = \left(\frac{1}{n}\right) \sum_{\tilde{E}_{i,i+1}>0} 1 \tag{37}$$

$$\hat{p}_{n}(NEG) = \left(\frac{1}{n}\right) \sum_{\tilde{E}_{i,n}<0} 1 \tag{38}$$

$$\hat{p}_n(POS) = \left(\frac{1}{n}\right) \sum_{\tilde{E}_{i,n}>0} 1 \tag{39}$$

Equations 36 and 37 denote the proportion of next-phase projections that are over or under-predicted: $\hat{p}_{i+1}(NEG)$ is the proportion of projections that under-predicted and $\hat{p}_{i+1}(POS)$ is the proportion projections that over-predicted. Equations 38 and 39 denote the proportion of end-phase projections that are over or under-predicted: $\hat{p}_n(NEG)$ is the proportion of projections that under-predicted and $\hat{p}_n(POS)$ is the proportion of projections that over-predicted.

Confidence intervals for Equations 36, 37, 38, and 39 are based on the standard normal distribution error for proportions for a given significance level α [21].

$$100(1-\alpha)\%CI = \hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{40}$$

The normality assumptions hold provided $n \geq 30$ and $np \geq 5$ [21]. This means that the confidence intervals do not hold for proportions near 1 or 0. While $n \geq 30$ for all models, there were cases where $np < 5$. In these instances, it has been shown by [23] that Equation 41 provides an adequate confidence interval.

$$100(1-\alpha)\%CI = \left(\frac{1}{1+\frac{1}{n}z^2}\right)\left[\hat{p} + \left(\frac{1}{2n}\right)z^2 \pm z\sqrt{\left(\frac{1}{n}\right)\hat{p}(1-\hat{p}) + \left(\frac{1}{4n^2}\right)z^2}\right] \tag{41}$$

where $z = z_{1-\alpha/2}$.

4.3.2 Response Means.

Once the max, final, average, and initial errors are calculated for each model, an Analysis of Variance (ANOVA) was conducted with the null hypothesis that each reliability growth projection model mean error was the same. If the ANOVA indicated

at least one of the models had a different mean, a pairwise comparison was conducted on the means via Tukey's Honestly Significant Difference (HSD) method (also known as the Tukey-Kramer Test). Tukey's HSD is a procedure that conducts pairwise comparisons of means in order to determine a difference at and overall significance level α. It is based on the distribution of the studentized range statistice, q.

$$q = \frac{\bar{y}_{max} - \bar{y}_{min}}{\sqrt{MSE/n}} \tag{42}$$

with $\bar{y}_{max}$ and $\bar{y}_{min}$ the largest and smallest estimated means[21]. The test statistic t_α is based on the q distribution, α-level, r, the number of comparisons, and f, the degrees of freedom for the MSE. A pair of estimated means ($\bar{y}_A$, $\bar{y}_B$) is considered significantly different if the absolute difference in means is greater than t_α[21].

$$t_\alpha = q_\alpha(r, f)\sqrt{\frac{MSE}{n}} \tag{43}$$

4.4 Model Results

4.4.1 Projection Proportions.

The single-phase projection errors are shown in the Appendix. The proportions and confidence intervals are in Tables 4 and 5.

Table 4. Proportions of Under-and-Over-Prediction onto Next Phase with Confidence Intervals

Run	Duane		VSD		ACPM		AMPM		AMPM-Stein		Clark	
	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over
Proportion	0.900	0.100	0.751	0.249	0.129	0.871	0.502	0.498	0.137	0.863	0.000	1.000
Standard Error	0.021	0.021	0.030	0.030	0.021	0.021	0.031	0.031	0.021	0.021	0.003	0.003
95% Lower Bound	0.859	0.060	0.693	0.190	0.089	0.830	0.441	0.438	0.095	0.822	-0.001	0.984
95%Upper Bound	0.940	0.141	0.810	0.307	0.170	0.911	0.562	0.559	0.178	0.905	0.009	0.994

In order to illustrate the magnitude of the over and under-prediction error, his-

Table 5. Proportions of Under-and-Over-Prediction onto Final Phase with Confidence Intervals

Run	Duane		VSD		ACPM		AMPM		AMPM-Stein		Clark	
	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over
Proportion	0.861	0.139	0.622	0.378	0.179	0.821	0.627	0.373	0.163	0.837	0.042	0.958
Standard Error	0.024	0.024	0.034	0.034	0.024	0.024	0.030	0.030	0.023	0.023	0.012	0.012
95% Lower Bound	0.814	0.092	0.556	0.312	0.132	0.775	0.569	0.314	0.119	0.792	0.018	0.934
95%Upper Bound	0.908	0.186	0.688	0.444	0.225	0.868	0.686	0.431	0.208	0.881	0.066	0.982

tograms of the errors were created (see Figure 2). It is worth noting that the Duane and Variance-Stabilized Duane models under-predict the increase in MTBF a significantly higher proportion of the time for both projections, while the ACPM, AMPM-Stein, and Clark models over-predict the increase in MTBF a significantly higher proportion of the time. In fact, the Clark model never under-predicted the next-phase MTBF in any run. The AMPM model prediction proportions were approximately 0.5 (not statistically different at the $\alpha = 0.05$ significance level), meaning that the AMPM over-predicts and under-predicts approximately the same proportion.

Based on the Figure 2, we can see that whenever a model over or under-predicts the MTBF increase, the projection is close to the observed MTBF, especially for the AMPM, Duane, and Variance-Stabilized Duane models. The ACPM, AMPM-Stein, and Clark models have higher errors, as we will show in 4.4.2.

4.4.2 Response Means.

For the ACPM, AMPM, AMPM-Stein, and Clark models, 54 runs were conducted while 270 were conducted for the Duane and Variance-Stabilized Duane models, due to the additional initial MTBF factor, resulting in a total of 3672 individual projections. Initial summary statistics are in Tables 6 through 9.

The Clark confidence intervals contain 0 for all response factors which indicates that the variance for those responses is high. Additionally, the confidence for max

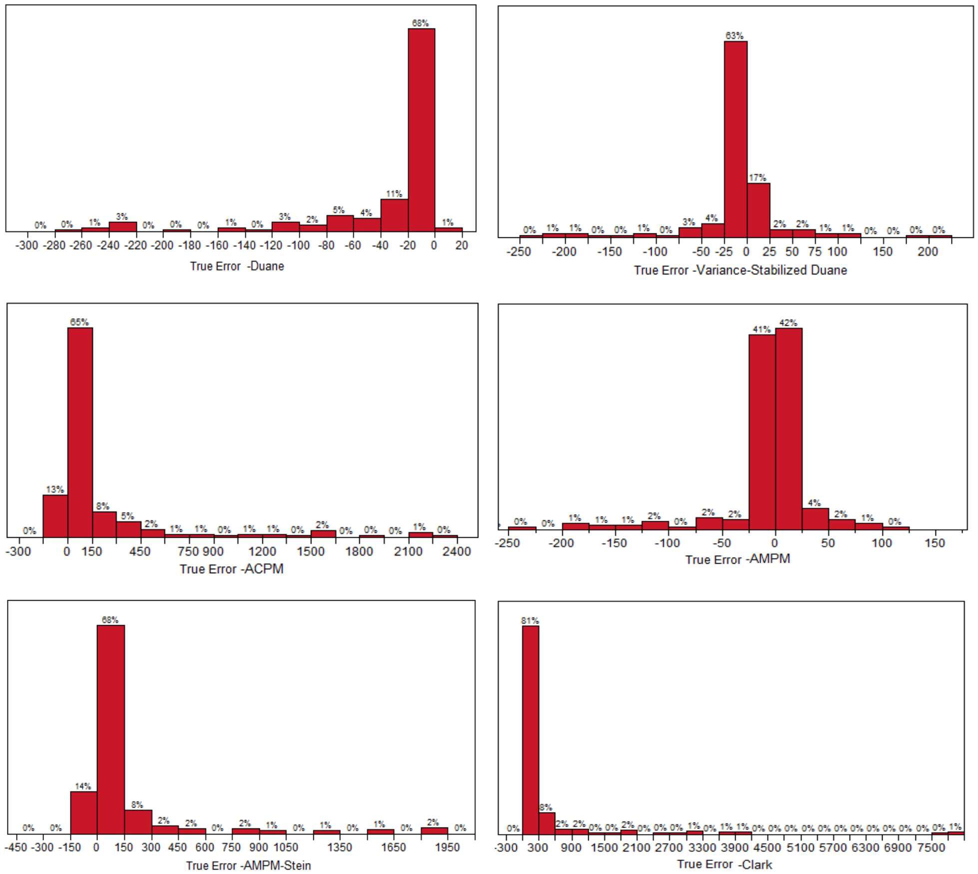


Figure 2. Histograms of True Projection Errors for Next Phase Projection

error, final error, and average error of the Clark model completely contain the confidence intervals for the other models. For all models, the max error is significantly higher than the average error, however, for the Clark model the max error is orders of magnitude higher. While this indicates that the Clark model has higher projection error, it also indicates that some runs may skew the data.

It was noted that the AMPM had the lowest average for the max, final, and average error. The AMPM-Stein and ACPM had the lowest errors for the projection from the initial phase to the end of test. While the Duane and Variance-Stabilized Duane models performed comparatively well in the average, max, and final phase errors, they had the highest errors (along with the AMPM) for the initial phase projections.

Table 6. Maximum Normalized Projection Error Summary Statistics

Model	Average	95% Lower Bound	95% Upper Bound	Std Dev	Max	Min
Duane	923.97	422.17	1425.76	1881.38	8055.65	0.00025
Variance-Stabilized Duane	701.28	320.73	1081.82	1426.77	4889.37	0.0012
ACPM	7078.84	3864.99	10292.69	12049.66	41204.94	1.65
AMPM	62.12	30.54	93.70	118.40	612.93	0.0014
AMPM-Stein	8180.30	2374.39	13986.20	21768.02	120125.00	1.97
Clark	60689.25	-1090.36	122468.85	231629.53	1632153.85	15.74

Table 7. Normalized Projection Error Summary Statistics (From Final Phase to End of Test)

Model	Average	95% Lower Bound	95% Upper Bound	Std Dev	Max	Min
Duane	229.91	-63.22	523.03	1099.01	8055.65	0.00025
Variance-Stabilized Duane	298.13	76.88	519.38	829.52	4290.87	0.0012
ACPM	2442.89	558.19	4327.58	7000.52	36628.34	1.65
AMPM	10.27	3.73	16.81	24.51	125.04	0.0000056
AMPM-Stein	6460.16	763.51	12156.81	21358.37	120125.00	1.97
Clark	53654.96	-8166.66	115476.57	231787.02	1632153.85	15.74

Table 8. Average Normalized Projection Error Summary Statistics

Model	Average	95% Lower Bound	95% Upper Bound	Std Dev	Max	Min
Duane	236.90	114.34	359.45	459.49	1892.57	0.00025
Variance-Stabilized Duane	191.98	88.17	295.79	389.21	1650.72	0.0012
ACPM	1508.15	831.07	2185.23	2538.57	9917.08	0.83
AMPM	14.22	8.23	20.21	22.45	88.03	0.00071
AMPM-Stein	1717.30	583.86	2850.74	4249.59	20162.90	0.99
Clark	12714.38	-67.14	25495.90	47921.58	341249.08	8.41

Table 9. Normalized Projection Error Summary Statistics (From First Phase to End of Test)

Model	Average	95% Lower Bound	95% Upper Bound	Std Dev	Max	Min
Duane	439.63	28.23	851.02	1542.44	11106.49	0.00025
Variance-Stabilized Duane	353.85	-57.65	765.35	1542.85	11199.02	0.0012
ACPM	8.60	1.61	15.58	26.19	178.30	0.000002
AMPM	29.00	15.74	42.25	49.70	282.54	0.000005
AMPM-Stein	8.53	1.61	15.46	25.97	177.30	0.008
Clark	14.66	5.32	24.01	35.04	228.93	0.00026

ANOVA revealed at least one of the model means was different from the others. Tukey's Honestly Significant Difference (HSD) was used to determine the pairwise mean differences. The results of Tukey's HSD are in Table 10.

For the max error, final error, and average error, the Clark model was shown to have a significantly higher average, at least three times as high as the next highest average. This indicated that the high variance in the Clark model responses skewed the results of the Tukey's test statistic. The ANOVA was conducted again, this time excluding the Clark model responses, the results showed that at least one of the

Table 10. Tukey's HSD Results: Different Letter Indicates a Difference of Means

Model	Max Error	Final Error	Average Error	Initial Error
Duane	A_1	A_2	A_3	A_4
Variance-Stabilized Duane	A_1	A_2	A_3	A_4
ACPM	A_1	A_2	A_3	B_4
AMPM	A_1	A_2	A_3	B_4
AMPM-Stein	A_1	A_2	A_3	B_4
Clark	B_1	B_2	B_3	B_4

remaining model averages was different. A second Tukey-Kramer test was conducted, the results can be seen in Table 11.

Table 11. Tukey's HSD Results: Clark Model Excluded

Model	Max Error	Final Error	Average Error	Initial Error
Duane	A_1	A_2	A_3	A_4
Variance-Stabilized Duane	A_1	A_2	A_3	A_4
ACPM	B_1	AB_2	B_3	B_4
AMPM	A_1	A_2	A_3	B_4
AMPM-Stein	B_1	B_2	B_3	B_4

The results of the second Tukey's HSD test show that for max error and average error, the ACPM and AMPM-Stein models are significantly different than the remaining models. Both models have significantly higher means, indicating that these models have higher maximum and average projection errors across all of the runs. The AMPM-Stein model also had a significantly higher least squares mean for the final projection error, indicating that it is not as accurate as the other models in projecting the MTBF at the end of the test.

The ANOVA and Tukey-Kramer tests were conducted on subsets of the data as well. The results of the ANOVA and Tukey-Kramer tests were consistent for exponential and Weibull-distributed subsets.

The model with the lowest average error is the AMPM. This is true for all AMPM responses except the initial error, which is discussed later. This is significant because the AMPM has the lowest max error, final error, and average error across all of the

runs. When combined with the lack of over or under-prediction tendencies, this suggests that the AMPM may be the best suited model for projection onto to the next phase. The AMPM-Stein and ACPM had the lowest projection errors in projections from the initial phase to the end of test, suggesting that they may be more appropriate for multi-phase projection.

4.4.3 Observations.

The Clark model clearly had higher projection error for all responses except for the initial projections. This was due to the assumption that no new failure modes would be discovered (the $h(t)$ term in Equation 7). This assumption led to lower projected failure rates and higher projected MTBF.

Both the Duane and the Variance-Stabilized Duane models were tested against the initial MTBF assumptions. From the ANOVA, there was no statistical difference between the projection errors, regardless of the assumed initial MTBF. While the assumed initial MTBF does affect the projection, the effect on the projection error was small due to the transformation required for each model, meaning that both the Duane and Variance-Stabilized Duane models are very robust to incorrect assumptions regarding the initial MTBF.

Despite having consistently low average, max, and final projection errors, the AMPM did not perform well when estimating the final MTBF from the initial stage. This is due to the fact that the projections for the AMPM tend to plateau for phases beyond the phase directly following the current phase. For projections into phases that were two or more CAPs later, the projected increase in MTBF was very low. See Figure 3 for an example. Note how the projections beyond the next phase remain steady and level. This means that the AMPM is not an appropriate model for projecting the MTBF after testing from any phase except the final.

The AMPM and AMPM-Stein models both utilize the same process for esti-

Run 5: 20 FMs, 5 CAPs, Exponentially-Distributed Failure Times

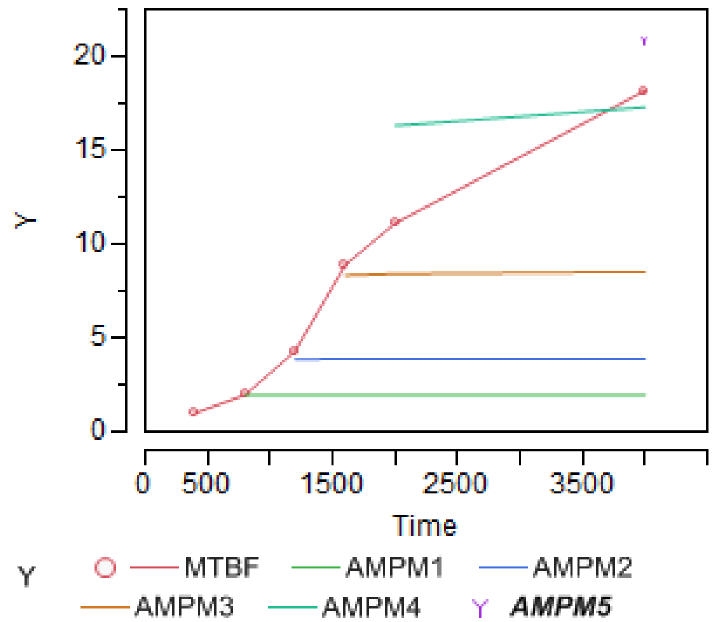


Figure 3. AMPM Projections onto All Subsequent Phases

mating the increase in reliability; however, this study suggests that the use of the Stein estimation process results in significantly increased error for next-phase projection. This may be due to the manner in which the Stein estimate parameter $\hat{\theta}_S$ is estimated. In this study, the Stein Estimate of the true failure rates were lower for phases with very few failures (less than 5). In such instances, the AMPM-Stein model tended towards over-prediction. However, the Stein-estimation process results in significantly lower error for longer-range projections (particularly projections from the initial phase onto the end of testing). This suggests that the Stein estimate performs well when there are more failures in a phase and has better projections into later testing times.

All models had increased projection errors for five of the six runs with 4 FMs and 8 CAPs (runs 7, 16, 25, 34, and 43). The five runs all had a significant deviation from

the other datasets: in at least one of the phases, no failures were observed. When this occurred, no fixes can take place, meaning that there is no reliability growth for any of the failure modes. The phase without failures is combined with the following phase, doubling the testing time. This deviation was only present when there were 4 FMs and 8 CAPs, presumably due to the shorter testing phases and relatively few FMs to observe. For the five runs in question, every model had higher projection error (the maximum projection error for all six models was in one of the five runs). Figure 4 shows the observed MTBF against the projected MTBF for all models. Note that the Duane, Variance-Stabilized Duane, and AMPM all under-predicted the reliability growth, while the ACPM, AMPM-Stein, and Clark models all over-predicted the growth (in the case of the Clark model, by a large margin).

Run 16: 4 FMs, 8 CAPs, Weibull-Distributed Failure Times

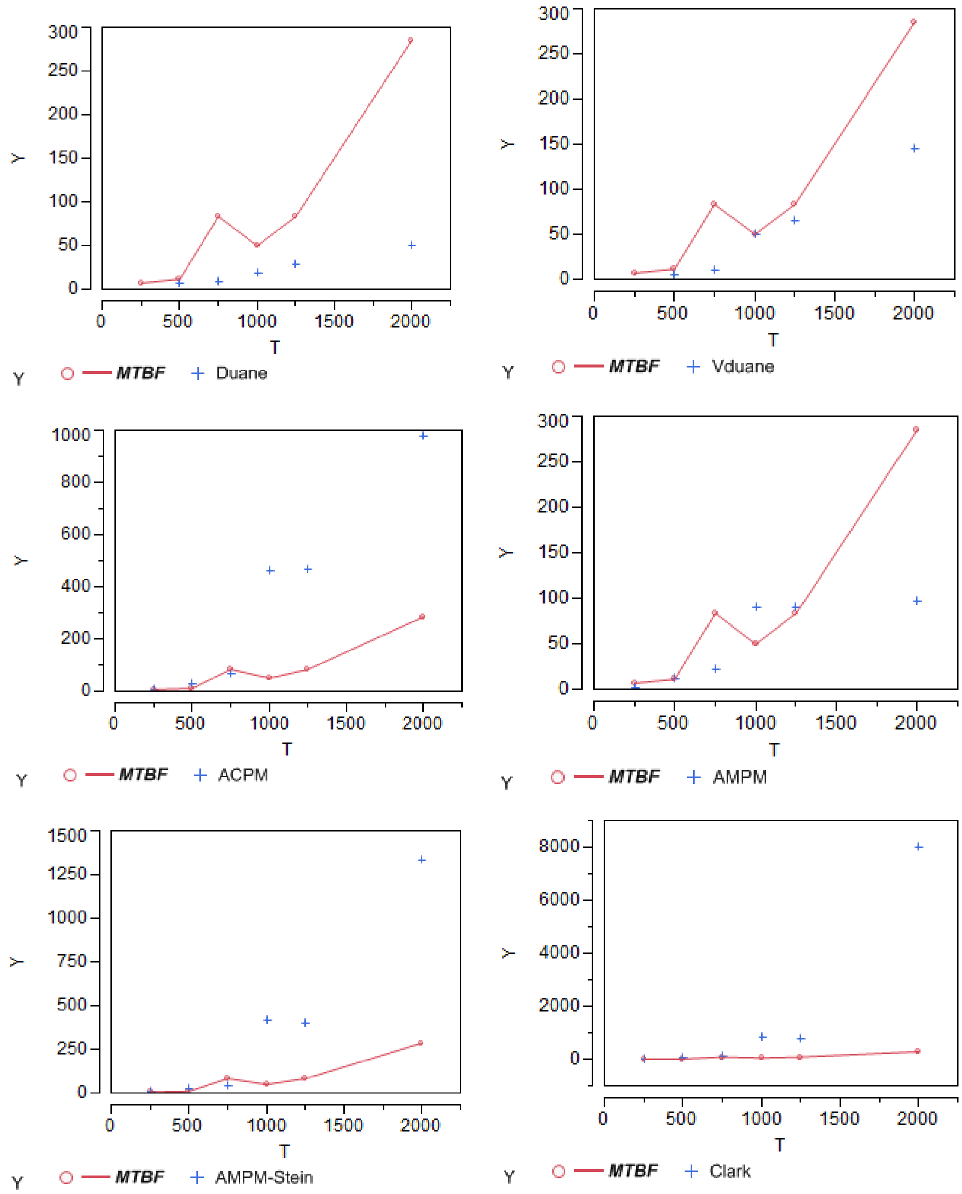


Figure 4. Model Projections vs Actual MTBF for All Models

4.5 Analysis Summary and Recommendations

Taking the results of the Tukey's tests and the proportions tests together, it would appear that the AMPM is significantly more accurate than all other models

for max error, final error, and average error. The fact that it does not tend towards over-prediction or under-prediction also indicates that the AMPM's performance is consistently robust to violations of the model assumptions, based on the results of this study.

Despite being the simplest of the models, the Duane and Variance-Stabilized Duane models still had significantly lower error in every response except the initial projection error. No significant difference could be determined between the Duane and the Variance-Stabilized Duane models, though the Variance-Stabilized Duane does have simpler calculations. Both the Duane and Variance-Stabilized Duane models tended to under-predict the MTBF, as seen in Tables 4 and 5. When considered together, the Duane and Variance-Stabilized Duane models provide a pessimistic estimate for the increase in reliability, based on the results of this study.

From the results of this study, all models that tended towards over-prediction also tended to have higher projection error due to the fact that there is no upper-bound on the maximum projection error for over-prediction as there is for under-prediction (there cannot be negative reliability). Despite this, the ACPM had the lowest error for next-phase projection out of the models with over-prediction tendencies (ACPM, Clark, AMPM-Stein). This suggests that the ACPM can provide an optimistic projection of the increase in reliability.

As previously discussed, the results of this study showed that the AMPM is poorly suited to project beyond the subsequent phase, and should not be used to estimate beyond one phase, especially from the initial testing phase. Based on the results of this study the AMPM-Stein model had the most accurate projections for the final MTBF from the initial phase. Because no fixes have occurred in the initial phase, this phase generally has the highest number of observed failures, avoiding the issues where the AMPM-Stein model underestimates the true failure rate as noted before.

Based on the results of this study, the AMPM-Stein model is best suited to project the final MTBF from the initial phase.

V. Conclusions

5.1 Thesis Summary

Reliability continues to be a matter of concern for both government acquisition and commercial enterprises. Determining the potential future reliability for a system at the beginning and throughout development and managing reliability growth effectively can have significant impacts to the planning and programing decisions and costs. Despite this, systems within the Department of Defense have consistently failed to meet the reliability thresholds [19] [14], which can lead to increased maintenance burdens and costs, as well as safety issues to personnel. While reliability growth projection has the potential to assist with these problems, the research prior to this study is inconclusive that it is a suitable tool.

In this study, six reliability growth projection models (Duane, Variance-Stabilized Duane, AMSAA-Crow Projection, AMSAA-Maturity Projection, AMSAA-Maturity Projection with Stein Estimation, and Clark Models) were used to project the change in reliability for 54 separate data sets produced via reliability testing simulation. Each model attempted to project the change in system reliability, making different assumptions regarding the nature of the system failures. The models were compared on the projection error and the tendency to over or under-project.

The results of this study suggest that the AMPM model is best suited for estimating the increase in reliability for the next phase, but is poorly suited for any estimation beyond that phase. With that, the AMPM-Stein variation is better suited for projecting the final MTBF from the initial phase, which can be used to determine the viability of the system. The Duane, Variance-Stabilized Duane, and AMPM models proved to be the most robust to violations in their assumptions, suggesting that these models would be most appropriate for reliability growth projection.

5.2 Future Research

5.2.1 Real-World Data.

Ideal testing should involve real-world testing data. When this study was conducted, few historical reliability testing datasets were readily available, and those that were failed to contain the necessary data for a reliability-growth projection model study. Should reliability growth testing documentation improve and be made available, incorporating this data in to future simulations would be vital to understanding the differences in projection model performance. Additionally, this would provide the ability to test model performance against actual reliability growth.

5.2.2 Extending the Simulation.

Lacking any historical reliability growth data, future simulations should incorporate additional aspects of complex systems that would violate additional reliability growth projection model assumptions.

- All six models tested assumed that repairs returned the system to the system state prior to the failure, but this may be unrealistic depending on the type of system being repaired. Future simulations should incorporate imperfect repairs, meaning that the repair may not completely undo the damage caused by the failure

- Failures may impact other areas of the system: a failure in FM 1 may increase or decrease the likelihood of observing a failure in FM 2. Incorporating dependent FMs into future simulations would test the robustness of all models against the independent failure mode assumptions

- Developing a simulation that allowed for Type A FMs and corrective actions during the phase (Test-Fix-Test corrective action implementation strategy) may

highlight differences in the models as well as allow for comparison of additional models like the ACPM-Extended

- Incorporating additional test articles into the simulation would provide a more accurate estimation on the observed MTBF for each phase

5.2.3 New Reliability Growth Projection Model Practices.

The results of this study show the model tendencies towards over and under-projection. It may be possible to improve the overall projection accuracy by using multiple models to develop multiple projections. This also suggests that changing the model used based on the current model projections may increase the accuracy of the projection. For example, if the current model used is consistently over-projecting, it may indicate that changing to the Duane or Variance-Stabilized Duane models would provide more accurate projections. As noted previously in this study, the ACPM and Duane models can provide optimistic and pessimistic projections, respectively. Finally, all models considered in this study were designed to function as standalone processes with no input other than the observed failures. It may benefit the projection process to consider how these models compare with standard forecasting methods, potentially using standard forecasting methods integrated with the reliability growth projection models to improve projection accuracy.

Appendix A. Proportion Tables

Table 12. Over-and-Under-Prediction Counts for MTBF Projection onto Next Phase - Replicate 1

Run	Duane		VSD		ACPM		AMPM		AMPM-Stein		Clark	
	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over
1	0	1	0	1	0	2	0	2	0	2	0	2
2	1	0	1	0	1	1	0	2	1	1	0	2
3	1	0	0	1	0	2	0	2	0	2	0	2
4	3	1	3	1	0	5	2	3	0	5	0	5
5	4	0	4	0	1	4	3	2	0	5	0	5
6	4	0	4	0	0	5	2	3	0	5	0	5
7	6	0	3	3	2	5	3	4	2	5	0	7
8	5	2	5	2	1	7	6	2	1	7	0	8
9	6	1	4	3	1	7	4	4	1	7	0	8
10	1	0	0	1	0	2	0	2	0	2	0	2
11	1	0	0	1	1	1	1	1	0	2	0	2
12	1	0	1	0	1	1	0	2	0	2	0	2
13	4	0	2	2	2	3	3	2	2	3	0	5
14	4	0	4	0	1	4	3	2	1	4	0	5
15	4	0	4	0	1	4	4	1	1	4	0	5
16	5	0	5	0	1	6	4	2	1	5	0	6
17	7	0	5	2	0	8	4	4	0	8	0	8
18	7	0	6	1	0	8	5	3	1	7	0	8

Table 13. Over-and-Under-Prediction Counts for MTBF Projection onto Next Phase - Replicate 2

Run	Duane		VSD		ACPM		AMPM		AMPM-Stein		Clark	
	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over
1	1	0	0	1	1	1	2	0	1	1	0	2
2	1	0	1	0	1	1	2	0	0	2	0	2
3	1	0	0	1	0	2	0	2	0	2	0	2
4	4	0	4	0	1	4	2	3	1	4	0	5
5	4	0	4	0	1	4	2	3	1	4	0	5
6	4	0	4	0	1	4	3	2	1	4	0	5
7	5	1	4	2	0	7	4	3	0	7	0	7
8	5	2	5	2	0	8	3	5	0	8	0	8
9	5	2	4	3	1	7	5	3	2	6	0	8
10	1	0	0	1	0	2	0	2	0	2	0	2
11	1	0	1	0	0	2	0	2	0	2	0	2
12	1	0	0	1	0	2	0	2	0	2	0	2
13	4	0	3	1	0	5	1	4	1	4	0	5
14	4	0	3	1	0	5	1	4	1	4	0	5
15	4	0	4	0	1	4	3	2	1	4	0	5
16	5	0	4	1	1	5	4	2	1	5	0	6
17	7	0	6	1	0	8	3	5	1	7	0	8
18	5	2	4	3	1	7	5	3	1	7	0	8

Table 14. Over-and-Under-Prediction Counts for MTBF Projection onto Next Phase - Replicate 3

Run	Duane		VSD		ACPM		AMPM		AMPM-Stein		Clark	
	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over
1	1	0	1	0	1	1	2	0	0	2	0	2
2	1	0	1	0	0	2	0	2	0	2	0	2
3	1	0	0	1	0	2	1	1	0	2	0	2
4	3	1	3	1	0	5	2	3	0	5	0	5
5	4	0	4	0	1	4	2	3	1	4	0	5
6	4	0	4	0	1	4	3	2	1	4	0	5
7	6	0	3	3	3	4	5	2	3	4	0	7
8	6	1	5	2	0	8	3	5	1	7	0	8
9	6	1	5	2	0	8	6	2	0	8	0	8
10	1	0	1	0	1	1	1	1	0	2	0	2
11	0	1	0	1	0	2	0	2	0	2	0	2
12	1	0	1	0	0	2	0	2	0	2	0	2
13	4	0	4	0	0	5	3	2	1	4	0	5
14	4	0	4	0	1	4	2	3	1	4	0	5
15	4	0	4	0	1	4	3	2	1	4	0	5
16	6	1	5	2	1	6	5	3	1	7	0	8
17	5	2	5	2	1	7	5	3	2	6	0	8
18	5	2	5	2	1	7	5	3	1	7	0	8

Table 15. Over-and-Under-Prediction Counts for MTBF Projection onto Final Phase - Replicate 1

Run	Duane		VSD		ACPM		AMPM		AMPM-Stein		Clark	
	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over
1	0	1	0	1	0	2	1	1	0	2	0	2
2	1	0	1	0	1	1	0	2	1	1	0	2
3	1	0	0	1	0	2	0	2	0	2	0	2
4	1	3	0	4	1	4	3	2	1	4	0	5
5	4	0	4	0	1	4	4	1	1	4	0	5
6	4	0	4	0	1	4	3	2	1	4	0	5
7	6	0	3	3	1	6	4	3	1	6	0	7
8	4	3	2	5	0	8	5	3	0	8	0	8
9	7	0	3	4	1	7	5	3	1	7	0	8
10	1	0	0	1	0	2	1	1	0	2	0	2
11	1	0	0	1	1	1	1	1	0	2	0	2
12	1	0	1	0	0	2	1	1	0	2	0	2
13	4	0	2	2	2	3	4	1	2	3	1	4
14	4	0	4	0	2	3	4	1	2	3	1	4
15	4	0	4	0	2	3	5	0	2	3	1	4
16	5	0	5	0	3	3	6	0	3	3	2	4
17	7	0	3	4	1	7	5	3	1	7	0	8
18	7	0	6	1	1	7	5	3	1	7	0	8

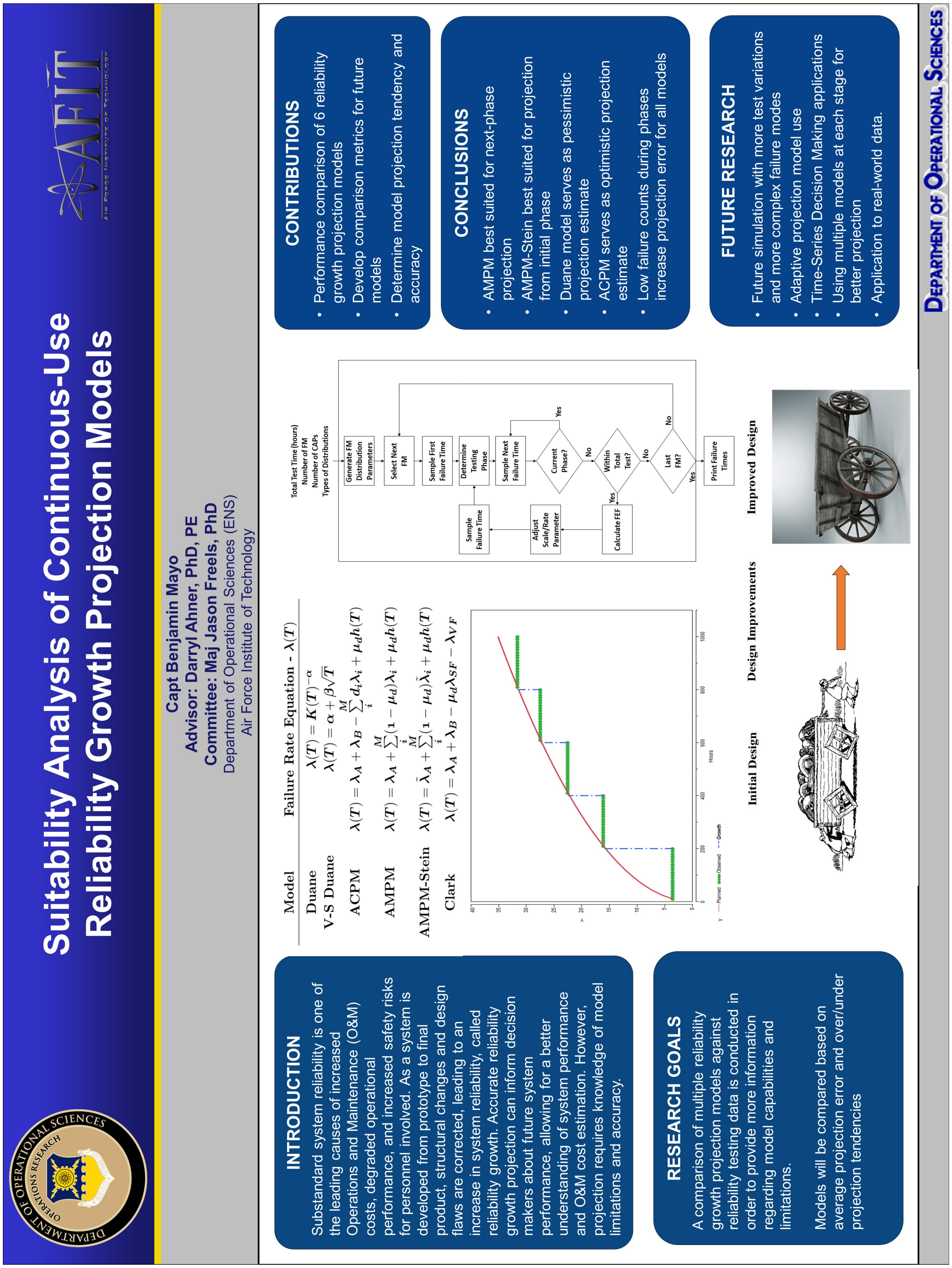
Table 16. Over-and-Under-Prediction Counts for MTBF Projection onto Final Phase - Replicate 2

Run	Duane		VSD		ACPM		AMPM		AMPM-Stein		Clark	
	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over
1	1	0	0	1	1	1	2	0	1	1	0	2
2	1	0	1	0	1	1	2	0	0	2	0	2
3	1	0	0	1	0	2	1	1	0	2	0	2
4	4	0	4	0	1	4	3	2	1	4	1	4
5	4	0	4	0	1	4	3	2	1	4	0	5
6	4	0	4	0	1	4	3	2	1	4	0	5
7	3	3	2	4	0	7	3	4	0	7	0	7
8	5	2	3	4	1	7	3	5	1	7	0	8
9	4	3	3	4	1	7	4	4	1	7	0	8
10	1	0	0	1	0	2	1	1	0	2	0	2
11	1	0	1	0	0	2	1	1	0	2	0	2
12	1	0	0	1	0	2	1	1	0	2	0	2
13	4	0	3	1	2	3	3	2	2	3	0	5
14	4	0	4	0	1	4	3	2	1	4	0	5
15	4	0	4	0	2	3	3	2	2	3	1	4
16	5	0	4	1	1	6	6	0	1	5	0	6
17	7	0	6	1	1	7	4	4	1	7	0	8
18	4	3	2	5	0	8	3	5	0	8	0	8

Table 17. Over-and-Under-Prediction Counts for MTBF Projection onto Final Phase - Replicate 3

Run	Duane		VSD		ACPM		AMPM		AMPM-Stein		Clark	
	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over	Under	Over
1	1	0	1	0	1	1	2	0	0	2	0	2
2	1	0	1	0	0	2	1	1	0	2	0	2
3	1	0	0	1	0	2	2	0	0	2	0	2
4	2	2	0	4	1	4	5	0	1	4	0	5
5	4	0	4	0	1	4	3	2	1	4	0	5
6	4	0	4	0	1	4	5	0	1	4	1	4
7	6	0	4	2	1	6	7	0	1	6	1	6
8	7	0	3	4	1	7	4	4	1	7	0	8
9	7	0	4	3	1	7	5	3	1	7	0	8
10	1	0	1	0	1	1	2	0	0	2	0	2
11	0	1	0	1	0	2	1	1	0	2	0	2
12	1	0	1	0	0	2	1	1	0	2	0	2
13	4	0	4	0	2	3	5	0	2	3	1	4
14	4	0	4	0	1	4	3	2	1	4	0	5
15	4	0	4	0	1	4	3	2	1	4	1	4
16	4	3	3	4	1	6	3	5	1	7	0	8
17	4	3	2	5	1	7	3	5	1	7	0	8
18	5	2	3	4	1	7	4	4	1	7	0	8

Appendix B. Storyboard



Bibliography

1. AMSC. *Department of Defense Handbook Reliability Growth Management.*

2. Jeffrey A. Clark. Modeling reliability growth late in development. *Reliability and Maintainability Symposium, 1999. Proceedings. Annual*, pages 201–207, Jan 1999.

3. National Research Council. *Reliability Issues for DOD Systems: Report of a Workshop.* The National Academies Press, 2002.

4. Larry H. Crow. Reliability analysis for complex, repairable systems. AMSAA Technical Report (138), 1975.

5. Larry H. Crow. Methods for assessing reliability growth potential. *Reliability and Maintainability Symposium, 1984. Proceedings. Annual*, pages 484–489, 1984.

6. Larry H. Crow. An extended reliability growth model for managing and assessing corrective actions. *Reliability and Maintainability, 2004 Annual Symposium - RAMS*, pages 73–80, Jan 2004.

7. John Donovan and Eamonn Murphy. Reliability growth - a new graphical model. *Quality and Reliability Engineering International*, 15(3):167–174, 1999.

8. John Donovan and Eamonn Murphy. A new reliability growth model: its mathematical comparison to the duane model. *Microelectronics Reliability*, 40(3):533–539, 2000.

9. J.T. Duane. Learning curve approach to reliability monitoring. *IEEE Transactions on Aerospace*, 2(2):563–566, April 1964.

10. Charles E. Ebeling. *An Introduction to Reliability and Maintainability Engineering.* Wavland Press, Inc, 2nd edition, 2010.

11. Paul M. Ellner and J. Brian Hall. AMSAA maturity projection model based on stein estimation. Technical report, DTIC Document, 2004.

12. Paul M Ellner and Lindalee C Wald. AMSAA maturity projection model. In *Reliability and Maintainability Symposium, 1995. Proceedings., Annual*, pages 174–181. IEEE, 1995.

13. Jason K. Freels. *Modeling Reliability Growth in Accelerated Testing.* PhD thesis, Air Force Institute of Technology, 2013.

14. Michael Gilmore. Key issues in reliability growth, September 2011. Presentation to National Academy of Science Panel on the Theory and Application of Reliability Growth Modeling in Defense Systems.

15. Huairui Guo, Wenbiao Zhao, and A. Mettas. Practical methods for modeling repairable systems with time trends and repair effects. *Reliability and Maintainability Symposium, 2006. RAMS '06. Annual*, pages 182–188, Jan 2006.

16. J. Brian Hall. *Methodology for Evaluating Reliability Growth Programs of Discrete Systems*. PhD thesis, University of Maryland, College Park, 2008.

17. John D. Kalbfleisch and Ross L. Prentice. *The Statistical Analysis of Failure Time Data*, volume 360. John Wiley & Sons, 2011.

18. William Q. Meeker and Luis A. Escobar. *Statistical Methods for Reliability Data*. Wiley-Interscience, 1st edition, 1998.

19. Panel on Reliability Growth Methods for Defense Systems. Reliability growth: Enhancing defense system reliability. Technical report, National Research Council, 2014.

20. Charles M. Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, pages 1135–1151, 1981.

21. Ronald E. Walphole. *Probability & Statistics for Engineers and Scientists*. Prentice Hall, 8th edition, 2006.

22. Herbert K. Weiss. Estimation of reliability growth in a complex system with a poisson-type failure. *Operations Research*, pages 532–545, 1956.

23. Edwin B. Wilson. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212, 1927.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
26–03–2015	Master's Thesis	October 2013 — March 2015

4. TITLE AND SUBTITLE

Suitability Analysis of Continuous-Use Reliability Growth Projection Models

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Mayo, Benjamin, R, Captain, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENS-MS-15-M-120

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Intentionally Left Blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution Statement A:
Approved For Public Release; Distribution Unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Substandard system reliability is one of the leading causes of increased Operations and Maintenance (O&M) costs as noted in several recent National Research Council reports. Between 2006 and 2011, Director Operational Test & Evaluation noted 26 of 52 Department of Defense acquisition programs failed to meet reliability thresholds, but were approved, leading to degraded operational performance, increased O&M costs, and increased safety risks for personnel involved. As a system is developed from prototype to final product, structural changes and design flaws are corrected, leading to an increase in system reliability, called reliability growth. Due to the nature of the system changes, standard forecasting methods cannot be applied, and a class of reliability growth models is used to estimate the change in reliability over multiple stages. Despite the significant impact of reliability growth projection, little research has been accomplished on comparing the robustness of various reliability growth models.

15. SUBJECT TERMS

Reliability Growth,Reliability Growth Projection,Reliability Simulation,Reliaibility Growth Model

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	65	Dr. Darryl Ahner, AFIT/ENS; (937) 255-3636, x4708; darryl.ahner@afit.edu